An Insight into Gill Microbiome of Eastern Mediterranean Wild Fish by Applying Next Generation Sequencing

Peleg Itay[a], Eli Shemesh[a], Maya Ofek-Lalzar[b], Nadav Davidovich[a,c], Yael Kroin[a], Shlomi Zrihan[a], Nir Stern[d], Arik Diamant[a], Natascha Wosnick[e] Dalit Meron[a], Dan Tchernov[a], Danny Morick[a,f,#]

[a]Morris Kahn Marine Research Station, Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Israel.

[b]Bioinformatics Services Unit, University of Haifa, Haifa, Israel.

[c]Israeli Veterinary Services, Bet Dagan, 5025001, Israel.

[d]National Institute of Oceanography, Israel Oceanographic and Limnological Research, Haifa, Israel

[e]Programa de Pós-graduação em Zoologia, Universidade Federal do Paraná, Brazil

[f]Hong Kong Branch of Southern Marine Science and Engineering, Guangdong Laboratory (Guangzhou), Hong Kong, China.

[#]Corresponding Author: Danny Morick (dmorick@univ.haifa.ac.il)

This PDF file includes:

Figures S1-S8

(Table S1 is an additional Excel file)

Supplementary methods

Supplementary figures

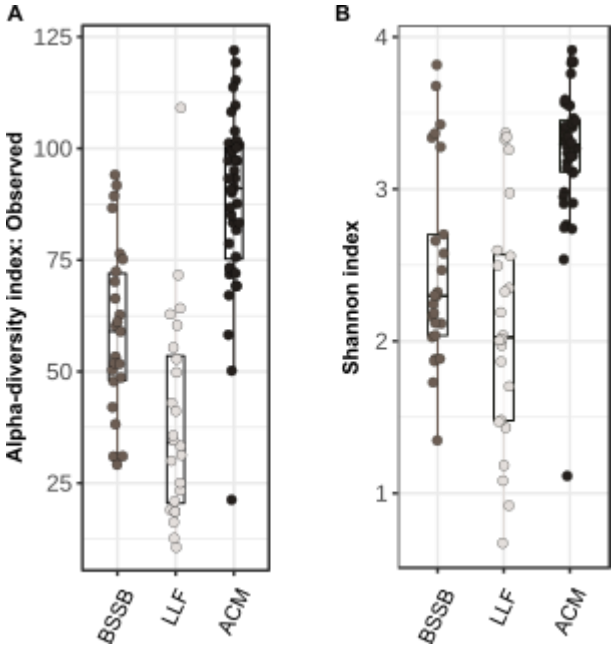

Figure S1: Alpha-diversity analyses comparison between fish samples using Observed (A) and Shannon index (B). Samples clustered by fish species. Observed: Kruskal-Wallis statistic: 44.181, *P*<.001. Shannon: Kruskal-Wallis statistic: 29.703, *P*<.001.

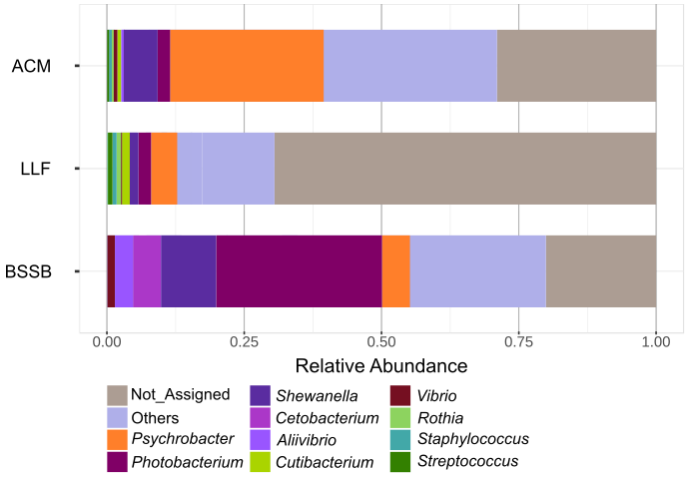

Figure S2: Average relative abundance of the fish species' major bacterial populations within the gills' community. The graph presents groups by relative size – from right to left – where the largest (more abundant) group is on the right and the least is on the left. The grey colored portions of the bars represent bacteria unidentified to the genus taxonomic level. Low count bacteria genera were aggregated into the 'Others' group. The three fish species express big differences in prominent genera, including presence of potential pathogens.

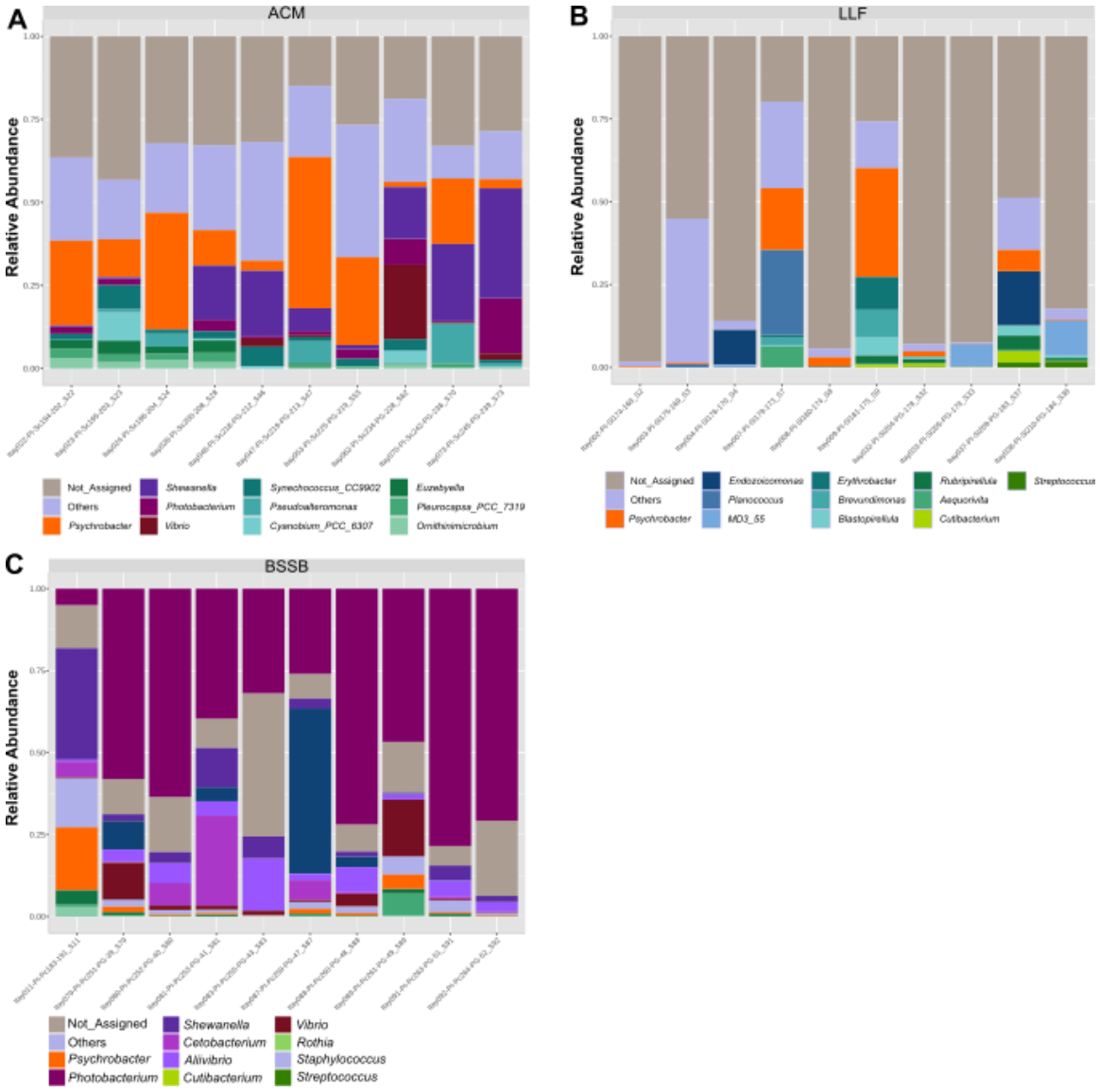

Figure S3: Atlantic Chub Mackerel (ACM; A), Lessepsian Lizardfish (LLF; B) and Bluespotted Seabream (BSSB; C) relative abundance – 10 random samples (each).

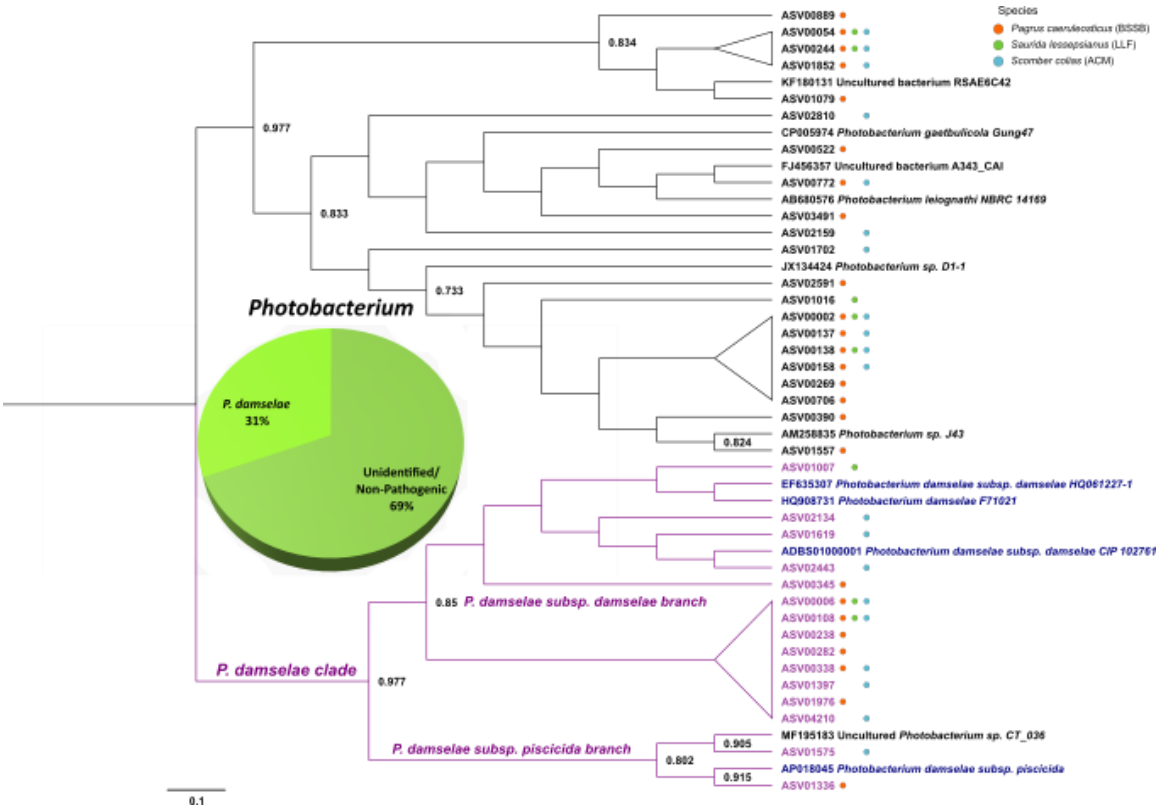

Figure S4: A phylogenetic tree for *Photobacterium*-related ASVs. A cutoff of 0.7 (70% bootstrap support) was made for nodes, thus any lower value is not presented. Triangular shaped tips represent sequences found to be practically identical. Grey-scale shapes (to the right of ASV numbers) represent fish species in which the ASVs appeared. Reference sequences include their GenBank accession numbers. Grey-colored sequence names mark pathogenic species and ASVs identified as bearing a similarity to pathogenic species-related sequences. The *P. damselae* clade branch and sub-branches are thickened, for emphasis. The pie chart refers to percentage out of total number of reads. Smaller ASV numbers indicate they were more common (in terms of total reads) than large number ASVs. The scale bar represents 0.1 nucleotide substitutions per site.

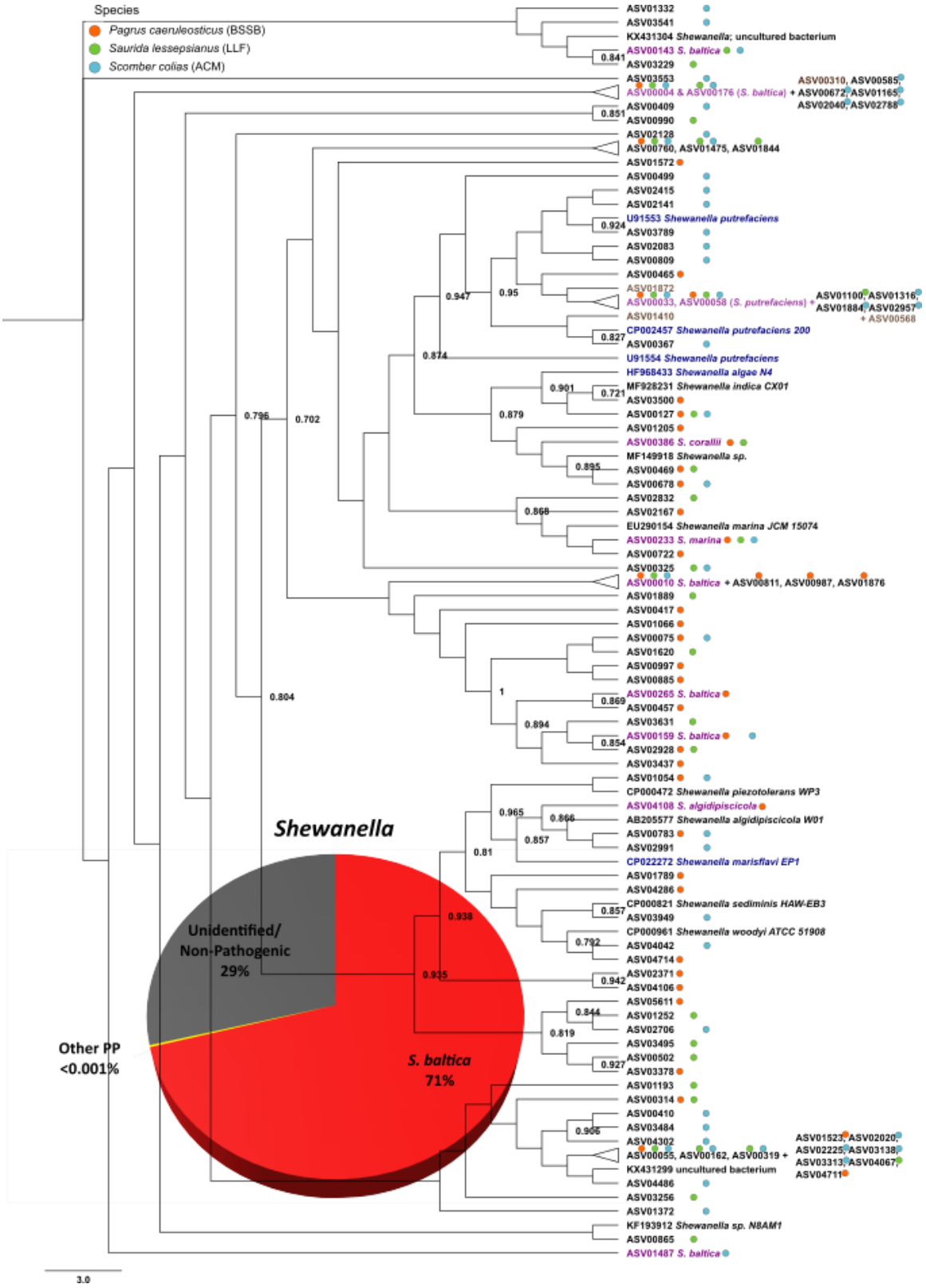
Species
Pagrus caeruleosticus (BSSB)
Saurida lessepsianus (LLF)
Scomber colias (ACM)
ASV01332
ASV03541
KX431304 Shewanella; uncultured bacterium
ASV00143 S. baltica
ASV03229
ASV03553
ASV00004 & ASV00176 (S. baltica) + ASV00310, ASV00585, ASV00672, ASV01165, ASV02040, ASV02788
ASV00409
ASV00990
ASV02128
ASV00760, ASV01475, ASV01844
ASV01572
ASV00499
ASV02415
ASV02141
U91553 Shewanella putrefaciens
ASV03789
ASV02083
ASV00809
ASV00465
ASV01872
ASV00033, ASV00058 (S. putrefaciens) + ASV01100, ASV01316, ASV01884, ASV02957 + ASV00568
ASV01410
CP002457 Shewanella putrefaciens 200
ASV00367
U91554 Shewanella putrefaciens
HF968433 Shewanella algae N4
MF928231 Shewanella indica CX01
ASV03500
ASV00127
ASV01205
ASV00386 S. corallii
MF149918 Shewanella sp.
ASV00469
ASV00678
ASV02832
ASV02167
EU290154 Shewanella marina JCM 15074
ASV00233 S. marina
ASV00722
ASV00325
ASV00010 S. baltica + ASV00811, ASV00987, ASV01876
ASV01889
ASV00417
ASV01066
ASV00075
ASV01620
ASV00997
ASV00885
ASV00265 S. baltica
ASV00457
ASV03631
ASV00159 S. baltica
ASV02928
ASV03437
ASV01054
CP000472 Shewanella piezotolerans WP3
ASV04108 S. algidipiscicola
AB205577 Shewanella algidipiscicola W01
ASV00783
ASV02991
CP022272 Shewanella marisflavi EP1
ASV01789
ASV04286
CP000821 Shewanella sediminis HAW-EB3
ASV03949
CP000961 Shewanella woodyi ATCC 51908
ASV04042
ASV04714
ASV02371
ASV04106
ASV05611
ASV01252
ASV02706
ASV03495
ASV00502
ASV03378
ASV01193
ASV00314
ASV00410
ASV03484
ASV04302
ASV00055, ASV00162, ASV00319 + ASV01523, ASV02020, ASV02225, ASV03138, ASV03313, ASV04067, ASV04711
KX431299 uncultured bacterium
ASV04486
ASV03256
ASV01372
KF193912 Shewanella sp. N8AM1
ASV00865
ASV01487 S. baltica
0.841
0.851
0.924
0.947
0.95
0.827
0.874
0.901
0.721
0.879
0.796
0.702
0.895
0.868
0.804
1
0.869
0.894
0.854
0.965
0.866
0.857
0.81
0.857
0.792
0.938
0.935
0.942
0.844
0.819
0.927
0.906
Shewanella
Unidentified/ Non-Pathogenic 29%
S. baltica 71%
Other PP <0.001%
3.0

Figure S5: A phylogenetic tree for Shewanella-related ASVs. A cutoff of 0.7 (70% bootstrap support) was made for nodes, thus any lower value is not presented. Triangular shaped tips represent sequences found to be practically identical. Grey-scale shapes (above or to the right of ASV numbers) represent fish species in which the ASVs appeared. Reference sequences include their GenBank accession numbers. Grey-colored sequence names mark pathogenic species and ASVs identified as bearing a similarity to pathogenic species-related sequences, while light-grey ASVs mark those found in negative control samples only. The pie chart refers to percentage out of total number of reads. Smaller ASV numbers indicate they were more common (in terms of total reads) than large number ASVs. The scale bar represents 3.0 nucleotide substitutions per site.

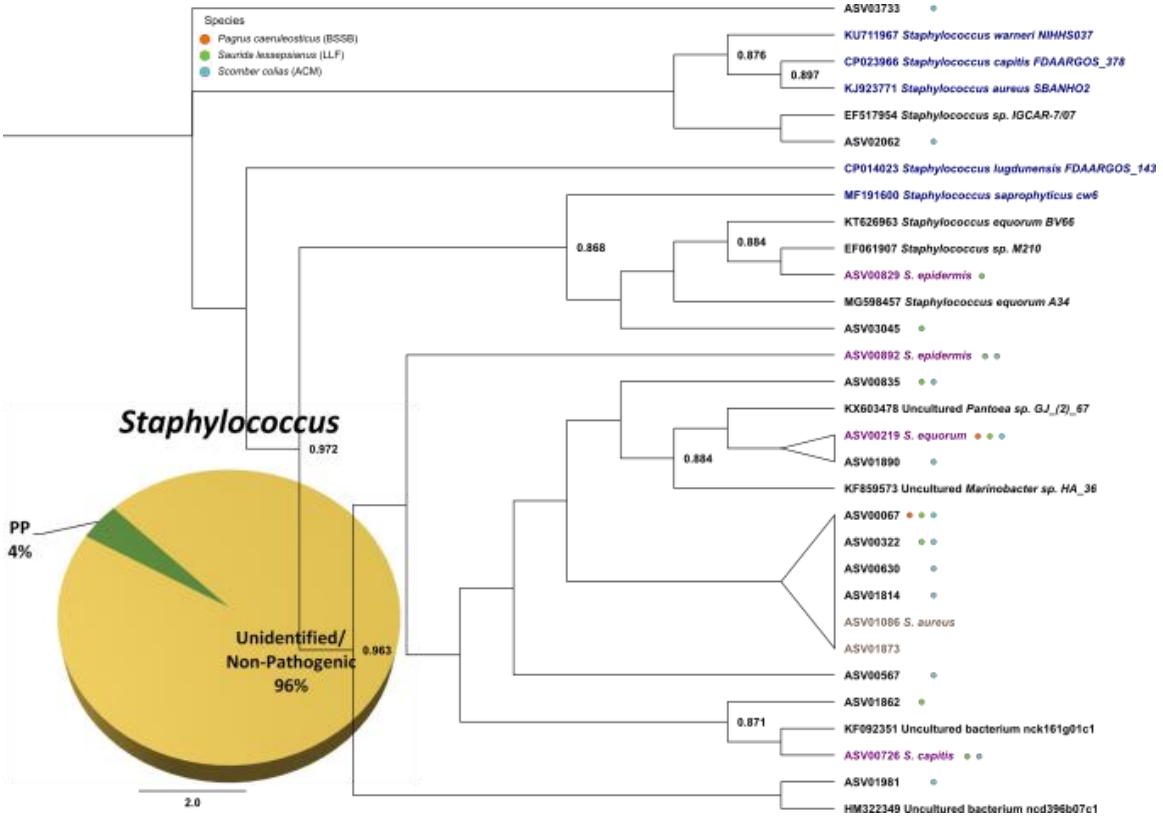

Figure S6: A phylogenetic tree for *Staphyloccocus*-related ASVs. A cutoff of 0.7 (70% bootstrap support) was made for nodes, thus any lower value is not presented. Triangular shaped tips represent sequences found to be practically identical. Grey-scale shapes (to the right of ASV numbers) represent fish species in which the ASVs appeared. Reference sequences include their GenBank accession numbers. Grey-colored sequence names mark pathogenic species and ASVs identified as bearing a similarity to pathogenic species-related sequences, while light-grey ASVs mark those found in negative control samples only. The pie chart refers to percentage out of total number of reads. Smaller ASV numbers indicate they were more common (in terms of total reads) than large number ASVs. The scale bar represents 2.0 nucleotide substitutions per site.

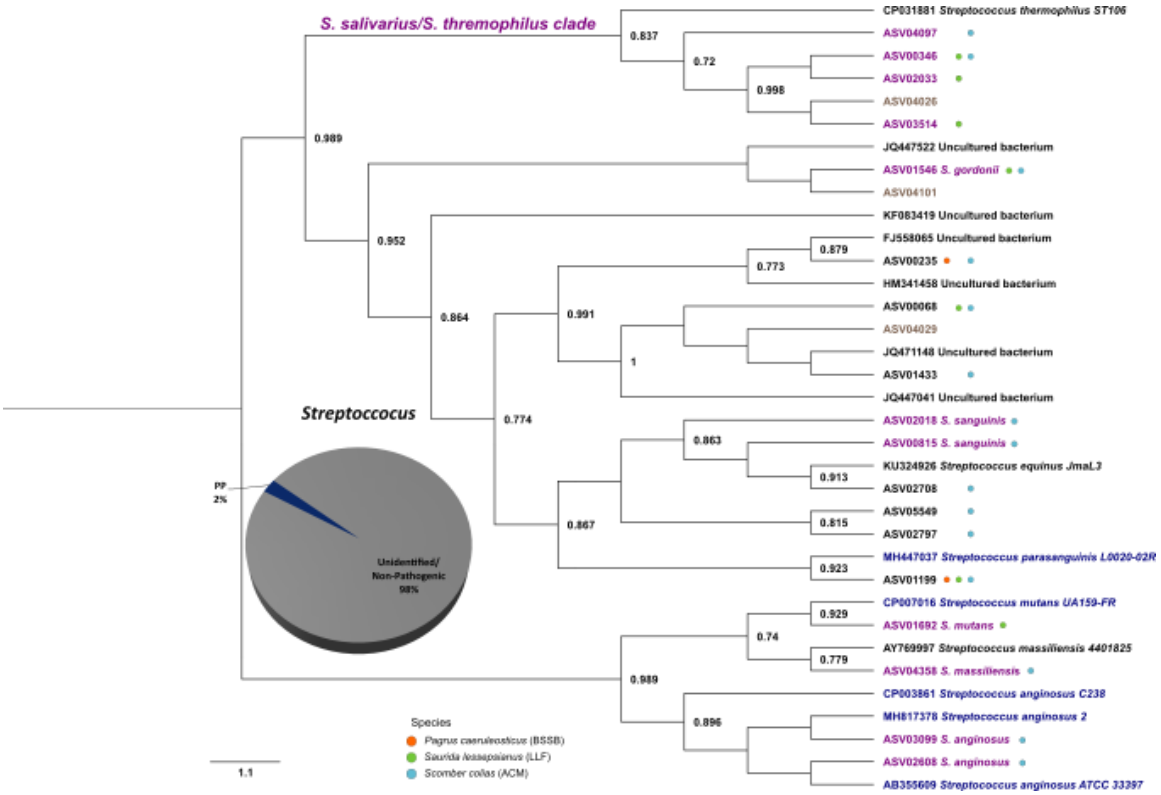

Figure S7: A phylogenetic tree for *Streptoccocus*-related ASVs. A cutoff of 0.7 (70% bootstrap support) was made for nodes, thus any lower value is not presented. Triangular shaped tips represent sequences found to be practically identical. Grey-scale shapes (to the right of ASV numbers) represent fish species in which the ASVs appeared. Reference sequences include their GenBank accession numbers. Grey-colored sequence names mark pathogenic species and ASVs identified as bearing a similarity to pathogenic species-related sequences, while light-grey ASVs mark those found in negative control samples only. The pie chart refers to percentage out of total number of reads. Smaller ASV numbers indicate they were more common (in terms of total reads) than large number ASVs. The scale bar represents 1.1 nucleotide substitutions per site.

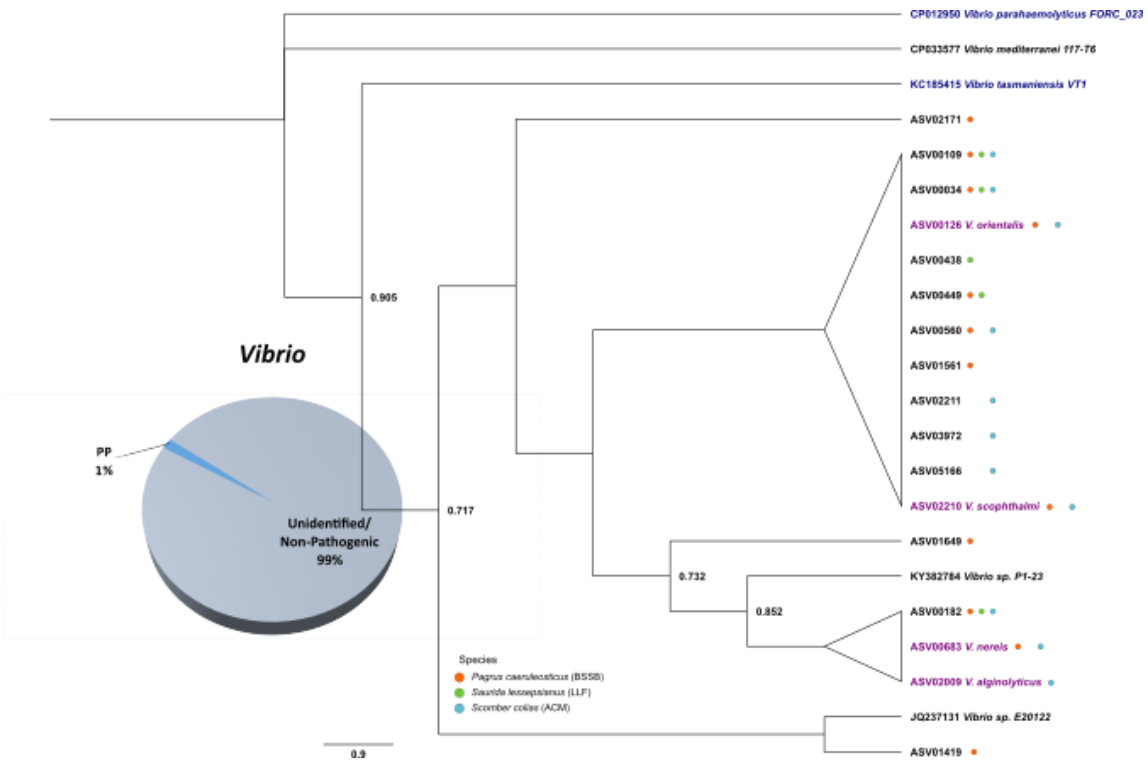

Figure S8: A phylogenetic tree for *Vibrio*-related ASVs. A cutoff of 0.7 (70% bootstrap support) was made for nodes, thus any lower value is not presented. Triangular shaped tips represent sequences found to be practically identical. Grey-scale shapes (to the right of ASV numbers) represent fish species in which the ASVs appeared. Reference sequences include their GenBank accession numbers. Grey-colored sequence names mark pathogenic species and ASVs identified as bearing a similarity to pathogenic species-related sequences. The pie chart refers to percentage out of total number of reads. Smaller ASV numbers indicate they were more common (in terms of total reads) than large number ASVs. The scale bar represents 0.9 nucleotide substitutions per site.

Supplementary Methods

Sequence data processing

Sequence data was analyzed using the Dada2 pipeline (72) using R package ‘dada2’ (version 1.14.1). Fastq formatted reads were trimmed and filtered for low quality using the command ‘filterAndTrim’ with parameters maxEE=2, maxN=0, trimleft=20 and the trunclen=150. Error rate estimation was carried out using the ‘learnerror’ command with default parameters, but with the randomize parameter set to TRUE, in order to sample nucleotides and reads for model

building randomly across all samples. The dada2 algorithm was implemented for error correction and a count table containing the amplicon sequence variants and counts per sample was produced. Merging of forward and reverse reads was done using the 'mergePairs' command with a minimum overlap of 8 bases. Then, suspected chimeras were detected and removed using the command 'removeBimeraDenovo', with default parameters. For each amplicon sequence variant (ASV), taxonomy (up to the species level) was inferred by alignment to the Silva non-redundant small subunit ribosomal RNA database (version 138), using commands 'assignTaxonomy' and 'addSpecies' with default parameters, while setting the minimum bootstrap confidence value to 80%.

Data analysis

For data analysis and generation of figures, the online tool MicrobiomeAnalyst (https://www.microbiomeanalyst.ca/MicrobiomeAnalyst/home.xhtml) was used (73,74). Through Marker Data Profiling (MDP), three CSV files were uploaded – ASV counts, taxonomy, metadata. Taxonomy labels were assigned using the SILVA taxonomic framework (https://www.arb-silva.de/documentation/silva-taxonomy/). Data filtering settings were defined as follows: (i) minimum counts: 4; (ii) low count filter – prevalence in samples (%): 10; (iii) low variance filter – percentage to remove (%): 5, and filter based on inter-quantile range. This process removed 1093 low abundance/variance features and kept 283 others. For Alpha-diversity graphs data normalization settings were: (i) data rarefied to the minimum library size; (ii) no data scaling; and (iii) no data transformation. Alpha-diversity profiling settings were: (i) data input – filtered; (ii) experimental factor – 'species'; (iii) taxonomic level – 'feature-level'; (iv) diversity measure – observed, Shannon and Simpson; and (v) the statistical method – Mann-Whitney/Kruskal-Wallis. For all other graphs and analyses, the data normalization settings required not rarefying data, but transforming the data using relative log expression (RLE). For

taxa abundance stacked-bar graphs, graph type was set to percentage-abundance, merging small taxa with counts < 8,000 and grouping by fish species. For beta-diversity graphs, method: NMDS ordination, distance method calculation: Jaccard index. Statistical method: Permutational MANOVA (PERMANOVA), other options kept as default. For the correlation analysis: algorithm – Spearman's rank correlation, *P*-value threshold – 0.05, correlation threshold – 0.5. All other values used default settings.

Phylogenetic trees

Sequences identified as belonging to the several genera chosen for deeper enquiry were copied from the main data file converted to FASTA format, and then uploaded to Silva (https://www.arb-silva.de/aligner/) for preparing phylogenetic files (75–77). The ACT (Alignment, Classification and Tree Service) tool was used (SINA v1.2.11) (ref. 68), with the following parameters: (i) gene: ssu; (ii) unaligned remaining bases – attached to the last aligned base; (iii) search and classify (checked) with minimum identity with query sequence set to 0.98, and number of neighbors per query sequence at 2; (iv) compute tree (checked) with its workflow set to 'denovo including neighbors', 'FastTree' as the program to use, and 'gamma' as the rate model for likelihoods; (v) the output format: FASTA and file zip-compressed; and (vi) taxonomies selected for classification: SILVA and RDP (only). All other parameters were kept as default. Output TREE format files were extracted for visualization with the FigTree v1.4.3 software (http://tree.bio.ed.ac.uk/software/figtree/).